ELSEVIER

Theoretical Computer Science 270 (2002) 947–958

Theoretical Computer Science

www.elsevier.com/locate/tcs

Note

An example of a computable absolutely normal number

Verónica Becher *, [1], Santiago Figueira

Departamento de Computatión, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina

Received January 2001; revised February 2001; accepted February 2001

Communicated by F. Cucker

Abstract

The first example of an absolutely normal number was given by Sierpinski in 1916, twenty years before the concept of computability was formalized. In this note we give a recursive reformulation of Sierpinski's construction which produces a computable absolutely normal number.

1. Introduction

"A number which is normal in any scale is called absolutely normal. The existence of absolutely normal numbers was proved by E. Borel. His proof is based on the measure theory and, being purely existential, it does not provide any method for constructing such a number. The first effective example of an absolutely normal number was given by me in the year 1916. As was proved by Borel almost all (in the sense of measure theory) real numbers are absolutely normal. However, as regards most of the commonly used numbers, we either know them not to be normal or we are unable to decide whether they are normal or not. For example we do not know whether the numbers $\sqrt{2}$, π, e are normal in the scale of 10. Therefore, though according to the theorem of Borel almost all numbers are absolutely normal, it was by no means easy to construct an example of an absolutely normal number. Examples of such numbers are fairly complicated."

(M.W. Sierpinski, 1964, p. 277.)

* Corresponding author.

E-mail addresses: vbecher@dc.uba.ar (V. Becher), sfigueir@ciudad.com.ar (S. Figueira).

[1] Supported by Agencia Nacional de Promoción Científica y Tecnológica, Argentina.

PII: S0304-3975(01)00170-0

A number is *normal* to base q if every sequence of n consecutive digits in its q-base expansion appears with limiting probability q^{-n}. A number is *absolutely normal* if it is normal to every base $q \geqslant 2$ [2]. For example, the rational number 0.1010101010... is not normal to base 2 because although the probability to find "1" is 2^{-1} and so is the probability to find "0", the probability to find "11" is not 2^{-2}. There are also irrational numbers that are not normal in some base, as 0.101001000100001000001..., which is not normal to base 2. Another example is Champernowne's number 0.12345678910111213 1415... which has all natural numbers in their natural order, written in base 10. It can be proved that Champernowne's number is normal to base 10, but not in some other bases. Let us notice that no rational is absolutely normal: a/b with $a<b$ is written in base b as $0.a000000000\ldots$ Moreover, every rational r is not normal to any base $q \geqslant 2$ [4]: the fractional expansion of r in base q will eventually repeat, say with a period of k, in which case the number r is about as far as being normal to the base q^k as it can be.

The first example of an absolutely normal number was given by Sierpinski in 1916 [5], twenty years before the concept of computability was formalized. Sierpinski determines such a number using a construction of infinite sets of intervals and using the minimum of an uncountable set. Thus, it is a priori unclear whether his number is computable or not. In this note we give a recursive reformulation of Sierpinski's construction which produces a computable absolutely normal number. We actually give an (ridiculously exponential) algorithm to compute this number.

The present work can be related to that of Turing [7], where he attempts to show how absolutely normal numbers may be constructed. However, the strategy he used is different from Sierpinski's and it is unclear whether the theorems announced in his paper actually hold.

Another example of an absolutely normal but not computable number is Chaitin's random number Ω, the halting probability of a universal machine [3]. Based on his theory of program size Chaitin formalizes the notion of lack of structure and unpredictability in the fractional expansion of a real number, obtaining a definition of randomness stronger than statistical properties of randomness. Although the definition of Ω is known there is no algorithm to exhibit its fractional digits. That is, Ω is not computable.

The fundamental constants, like π, $\sqrt{2}$ and e, are computable and it is widely conjectured [1, 6] that they are absolutely normal. However, none of these has even been proved to be normal to base 10, much less to all bases. The same has been conjectured of the irrational algebraic numbers [1]. In general, we lack an algorithm that decides on absolute normality.

Let us recall that a real number is computable if there is a recursive function that calculates each of its fractional digits. Namely, there exists a total recursive $f: N \to N$ such that for every n, $f(n)$ is the nth fractional digit of the number in some base.

We will use Lebesgue's definition of measure. The measure of the interval $I=(a,b)$, with $a<b$, is denoted by $\mu(I)=b-a$, and the measure of a set of intervals J, denoted

by $\mu(J)$ is the measure of $\bigcup J$. We will use several properties of Lebesgue's measure, e.g., countable additivity and subadditivity.

After presenting Sierpinski's original construction we introduce our algorithmic version of his construction. Then, we discuss Sierpinski's number and we consider other variants defining absolutely normal numbers.

2. Sierpinski's result of 1916

Sierpinski [5] achieves an elementary proof of an important proposition proved by Borel that states that almost all real numbers are absolutely normal. At the same time he gives way to effectively determine one such number. He defines $\Delta(\varepsilon)$ as a set of certain open intervals with rational end points. Although the set $\Delta(\varepsilon)$ contains countably many intervals, they do not cover the whole of the $(0,1)$ segment. Sierpinski proves that every real number in $(0,1)$ that is external to $\Delta(\varepsilon)$ is absolutely normal.

$\Delta(\varepsilon)$ is defined as the union of infinitely many sets of intervals $\Delta_{q,m,n,p}$. The parameter ε is a number in $(0,1]$ used to bound the measure of $\Delta(\varepsilon)$.

$$\Delta(\varepsilon) = \bigcup_{q=2}^{\infty} \bigcup_{m=1}^{\infty} \bigcup_{n=n_{m,q}(\varepsilon)}^{\infty} \bigcup_{p=0}^{q-1} \Delta_{q,m,n,p},$$

where q ranges over all possible bases, n ranges over the lengths of fractional expansions, p ranges between 0 and $q-1$, m allows for arbitrarily small differences in the rate of appearance of the digit p in the fractional expansions, and $\Delta_{q,m,n,p}$ is the set of all open intervals of the form $(b_1/q + b_2/q^2 + \cdots + b_n/q^n - 1/q^n, b_1/q + b_2/q^2 + \cdots + b_n/q^n + 2/q^n)$ such that $|c_p(b_1, b_2, \ldots, b_n)/n - 1/q| \geqslant 1/m$, where $0 \leqslant b_i \leqslant q-1$ for $1 \leqslant i \leqslant n$ and $c_p(b_1, b_2, \ldots, b_n)$ represents the number of times that the digit p appears amongst $b_1, b_2, \ldots, b_n$.

The idea is that $\Delta_{q,m,n,p}$ contains all numbers that are not normal to base q. If a number is normal in base q we expect that the rate of appearance of the digit p in a prefix of length n to be as close as possible to $1/q$. Each interval in $\Delta_{q,m,n,p}$ that contains all numbers written in base q start with $0.b_1 b_2 \ldots b_n$ and the digit p appears in $0.b_1 b_2 \ldots b_n$ at a rate different from $1/q$. Let us observe that the right end of the intervals in $\Delta_{q,m,n,p}$ add $2/q^n$: added only $1/q^n$ would leave the number $0.b_1 b_2 \ldots b_n 1111\ldots$ outside the open interval. Each interval in $\Delta_{q,m,n,p}$ has measure $3/q^n$ and for fixed q, m, n, $\Delta_{q,m,n,p}$ is a finite set.

From Sierpinski's proof follows that $n_{m,q}(\varepsilon)$ must be large enough as to imply $\mu(\Delta(\varepsilon)) < \varepsilon$; $n_{m,q}(\varepsilon) = \lfloor 24m^6 q^2/\varepsilon \rfloor + 2$ suffices. In order to bound the measure of $\Delta(\varepsilon)$, Sierpinski works with the sum of the measures of each interval of $\Delta_{q,m,n,p}$:

$$s(\varepsilon) = \sum_{q=2}^{\infty} \sum_{m=1}^{\infty} \sum_{n=n_{m,q}(\varepsilon)}^{\infty} \sum_{p=0}^{q-1} \sum_{I \in \Delta_{q,m,n,p}} \mu(I)$$

and he proves that $\mu(\Delta(\varepsilon)) \leqslant s(\varepsilon) < \varepsilon$ for every $\varepsilon \in (0,1]$.

Sierpinski defines $E(\varepsilon)$ as the set of all real numbers in $(0,1)$ external to every interval of $\Delta(\varepsilon)$ and he proves that for every $\varepsilon \in (0,1]$, every real in $E(\varepsilon)$ is absolutely normal. Since $\mu(E(\varepsilon))$ is greater than or equal to $(1-\varepsilon)$, for every ε in $(0,1]$, every real in $(0,1)$ is absolutely normal with probability 1.

Although the measure of $\Delta(\varepsilon)$ is less than ε, no segment (c,d) with $c<d$ can be completely included in $E(\varepsilon)$. If this happened there would be infinitely many rationals belonging to $E(\varepsilon)$, contradicting absolute normality. Thus, $E(\varepsilon)$ is a set of infinitely many isolated irrational points. Sierpinski defines $\xi = \min(E(1))$, and in this way he gives an effective determination of an absolutely normal number.

3. An algorithm to construct an absolutely normal number

Our work is based on one essential observation: we can give a recursive enumeration of Sierpinski's set $\Delta(\varepsilon)$, and we can bound the measure of error in each step. To simplify notation, we will fix a rational $\varepsilon \in (0, \frac{1}{2}]$ (in fact, ε can be any computable real in $(0, \frac{1}{2}]$) and we rename $\Delta = \Delta(\varepsilon)$; $s = s(\varepsilon)$; $n_{m,q} = n_{m,q}(\varepsilon)$. We define the computable sequence (Δ_k):

$$\Delta_k = \bigcup_{q=2}^{k+1} \bigcup_{m=1}^{k} \bigcup_{n=n_{m,q}}^{k \cdot n_{m,q}} \bigcup_{p=0}^{q-1} \Delta_{q,m,n,p}.$$

We also define the bounds to the measure of each term in the sequence:

$$s_k = \sum_{q=2}^{k+1} \sum_{m=1}^{k} \sum_{n=n_{m,q}}^{k \cdot n_{m,q}} \sum_{p=0}^{q-1} \sum_{I \in \Delta_{q,m,n,p}} \mu(I).$$

It is clear that $\lim_{k\to\infty} s_k = s$ and $\lim_{k\to\infty} \Delta_k = \Delta$. Since s_k is the sum of the measures of all intervals belonging to Δ_k, we have $\mu(\Delta_k) \leqslant s_k$ and similarly we have $\mu(\Delta) \leqslant s$. Besides, for every natural k, $s_k \leqslant s$. Let us observe that for every pair of natural numbers k and l such that $k \leqslant l$, we have $\Delta_k \subseteq \Delta_l$, and for any k, $\Delta_k \subseteq \Delta$. Finally, we define the error of approximating s by s_k, $r_k = s - s_k$. We can give a bound on r_k, a result that makes our construction computable.

Theorem 1. *For every natural number k, $r_k < 5\varepsilon/2k$.*

Proof. Let us define $S_{qmn} = \sum_{p=0}^{q-1} \sum_{I \in \Delta_{q,m,n,p}} \mu(I)$. By splitting the sums in the definition of s,

$$s = \sum_{q=2}^{k+1} \sum_{m=1}^{k} \sum_{n=n_{m,q}}^{k \cdot n_{m,q}} S_{qmn} + \sum_{q=2}^{k+1} \sum_{m=1}^{k} \sum_{n=k \cdot n_{m,q}+1}^{\infty} S_{qmn} + \sum_{q=2}^{k+1} \sum_{m=k+1}^{\infty} \sum_{n=n_{m,q}}^{\infty} S_{qmn}$$
$$+ \sum_{q=k+2}^{\infty} \sum_{m=1}^{\infty} \sum_{n=n_{m,q}}^{\infty} S_{qmn}. \tag{1}$$

But the first term of (1) is s_k, so the rest is r_k.

$$r_k = \sum_{q=2}^{k+1} \sum_{m=1}^{k} \sum_{n=k\cdot n_{m,q}+1}^{\infty} S_{qmn} + \sum_{q=2}^{k+1} \sum_{m=k+1}^{\infty} \sum_{n=n_{m,q}}^{\infty} S_{qmn} + \sum_{q=k+2}^{\infty} \sum_{m=1}^{\infty} \sum_{n=n_{m,q}}^{\infty} S_{qmn}. \quad (2)$$

We will bound each of the three terms that appear in this equation. From Sierpinski's proof we know that $S_{qmn} < 12m^4/n^2$. The third term of Eq. (2) can be bounded by

$$\sum_{q=k+2}^{\infty} \sum_{m=1}^{\infty} \sum_{n=n_{m,q}}^{\infty} S_{qmn} < 12 \sum_{q=k+2}^{\infty} \sum_{m=1}^{\infty} \left(m^4 \sum_{n=n_{m,q}}^{\infty} \frac{1}{n^2} \right). \quad (3)$$

Let us now find a bound for $\sum_{n=n_{m,q}}^{\infty} 1/n^2$. For any $i \geqslant 1$, it holds that $\sum_{n=i+1}^{\infty} 1/n^2 < 1/i$, and by the definition of $n_{m,q}$ we have, $n_{m,q} - 1 = \lfloor 24m^6q^2/\varepsilon \rfloor + 1 > 24m^6q^2/\varepsilon$. Then, $\sum_{n=n_{m,q}}^{\infty} 1/n^2 < \varepsilon/24m^6q^2$.

Applying this last result to (3) we have

$$\sum_{q=k+2}^{\infty} \sum_{m=1}^{\infty} \sum_{n=n_{m,q}}^{\infty} S_{qmn} < \frac{\varepsilon}{2} \left(\sum_{q=k+2}^{\infty} \frac{1}{q^2} \right) \left(\sum_{m=1}^{\infty} \frac{1}{m^2} \right) < \frac{\varepsilon}{k+1} < \frac{\varepsilon}{k}. \quad (4)$$

Similarly, the second term of Eq. (2) can be bounded by

$$\sum_{q=2}^{k+1} \sum_{m=k+1}^{\infty} \sum_{n=n_{m,q}}^{\infty} S_{qmn} < \frac{\varepsilon}{2} \left(\sum_{q=2}^{\infty} \frac{1}{q^2} \right) \left(\sum_{m=k+1}^{\infty} \frac{1}{m^2} \right) < \frac{\varepsilon}{2k}. \quad (5)$$

Finally, the first term of Eq. (2) can be bounded by

$$\sum_{q=2}^{k+1} \sum_{m=1}^{k} \sum_{n=k\cdot n_{m,q}+1}^{\infty} S_{qmn} < 12 \sum_{q=2}^{\infty} \sum_{m=1}^{\infty} \left(m^4 \sum_{n=k\cdot n_{m,q}+1}^{\infty} \frac{1}{n^2} \right) < \frac{\varepsilon}{k}. \quad (6)$$

Replacing in (2) the bounds found in (4)–(6) we conclude $r_k < \varepsilon/k + \varepsilon/2k + \varepsilon/k = 5\varepsilon/2k$. $\square$

We now give an overview of our construction of the binary number $v = 0.b_1\, b_2\, b_3 \ldots$. To determine the first digit of v we divide the $[0,1]$ interval in two halves, $c_0^1 = [0, \frac{1}{2}]$ and $c_1^1 = [\frac{1}{2}, 1]$, each of measure $\frac{1}{2}$. Thinking in base 2, in c_0^1 there are only numbers whose first fractional digit is 0 while in c_1^1 there are only numbers whose first fractional digit is 1. By Sierpinski's result we know that neither Δ nor any of the Δ_k cover the whole segment $[0,1]$. Of course, all points external to Δ must either be in c_0^1 or in c_1^1. The idea now is to determine a subset of Δ, Δ_{p_1}, big enough (i.e. sufficiently similar to Δ) as to ensure that, whenever Δ_{p_1} does not cover completely a given interval, then Δ does not either. We can guarantee this because we have an upper bound on the error of approximating Δ at every step. We pick the interval c_0^1 or c_1^1, the least covered by Δ_{p_1}. If we select c_0^1 then there will be real numbers external to every interval of Δ whose first digit in the binary expansion is 0; therefore, we define $b_1 = 0$. Similarly, if we select c_1^1 we define $b_1 = 1$.

To define the rest of the digits we proceed recursively. We will divide $c_{b_{n-1}}^{n-1}$ in two halves defining the intervals c_0^n and c_1^n, each of measure $1/2^n$. At least one of the two will not be completely covered by Δ. The nth digit of v will be determined by comparing the measure of a suitable set Δ_{p_n} restricted to the intervals c_0^n and c_1^n, where the index p_n is obtained computably from n. If we select c_0^n, then we will define b_n to be 0, otherwise b_n will be 1. Since these measures are computable we have obtained an algorithm to define a real number v, digit by digit, such that v is external to every interval of Δ. By Sierpinski's result v is absolutely normal.

Before proceeding we shall prove some results. The following proposition gives us a bound on the measure of the sets that have not been enumerated in the step k.

Proposition 2. *For every k, $\mu(\Delta - \Delta_k) \leqslant r_k$.*

Proof. Since Δ_k is included in Δ, the measure of $\Delta - \Delta_k$ is less than or equal to the sum of the measures of those intervals in Δ but not in Δ_k. Hence

$$\begin{aligned}\mu(\Delta - \Delta_k) &\leqslant \sum_{q=2}^{\infty}\sum_{m=1}^{\infty}\sum_{n=n_{m,q}}^{\infty}\sum_{p=0}^{q-1}\sum_{\substack{I \in \Delta_{q,m,n,p}\\ I \notin \Delta_k}} \mu(I)\\ &\leqslant s - \sum_{q=2}^{\infty}\sum_{m=1}^{\infty}\sum_{n=n_{m,q}}^{\infty}\sum_{p=0}^{q-1}\sum_{\substack{I \in \Delta_{q,m,n,p}\\ I \in \Delta_k}} \mu(I).\end{aligned}$$

But

$$\begin{aligned}\sum_{q=2}^{\infty}\sum_{m=1}^{\infty}\sum_{n=n_{m,q}}^{\infty}\sum_{p=0}^{q-1}\sum_{\substack{I \in \Delta_{q,m,n,p}\\ I \in \Delta_k}} \mu(I) &\geqslant \sum_{q=2}^{k+1}\sum_{m=1}^{k}\sum_{n=n_{m,q}}^{k\cdot n_{m,q}}\sum_{p=0}^{q-1}\sum_{\substack{I \in \Delta_{q,m,n,p}\\ I \in \Delta_k}} \mu(I)\\ &= \sum_{q=2}^{k+1}\sum_{m=1}^{k}\sum_{n=n_{m,q}}^{k\cdot n_{m,q}}\sum_{p=0}^{q-1}\sum_{I \in \Delta_{q,m,n,p}} \mu(I) = s_k.\end{aligned}$$

Hence, $\mu(\Delta - \Delta_k) \leqslant s - s_k = r_k$. $\square$

We are also able to bound the measure of the difference between two sets enumerated in different steps.

Proposition 3. *For any natural numbers k and l such that $k \leqslant l$, $\mu(\Delta_l - \Delta_k) \leqslant r_k - r_l$.*

Proof.

$$\mu(\Delta_l - \Delta_k) \leqslant \sum_{q=2}^{l+1} \sum_{m=1}^{l} \sum_{n=n_{m,q}}^{l\cdot n_{m,q}} \sum_{p=0}^{q-1} \sum_{\substack{I\in\Delta_{q,m,n,p}\\ I\notin\Delta_k}} \mu(I) = s_l - \sum_{q=2}^{l+1} \sum_{m=1}^{l} \sum_{n=n_{m,q}}^{l\cdot n_{m,q}} \sum_{p=0}^{q-1} \sum_{\substack{I\in\Delta_{q,m,n,p}\\ I\in\Delta_k}} \mu(I)$$

and as

$$k \leqslant l, \sum_{q=2}^{l+1} \sum_{m=1}^{l} \sum_{n=n_{m,q}}^{l\cdot n_{m,q}} \sum_{p=0}^{q-1} \sum_{\substack{I\in\Delta_{q,m,n,p}\\ I\in\Delta_k}} \mu(I) \geqslant \sum_{q=2}^{k+1} \sum_{m=1}^{k} \sum_{n=n_{m,q}}^{k\cdot n_{m,q}} \sum_{p=0}^{q-1} \sum_{I\in\Delta_{q,m,n,p}} \mu(I) = s_k.$$

Thus, $\mu(\Delta_l - \Delta_k) \leqslant s_l - s_k = (s - s_k) - (s - s_l) = r_k - r_l$. □

Let J be a set of intervals and let c be an interval. We will denote with $J \cap c$ the restriction of J to c, $J \cap c = \{x \in \mathbb{R}\colon\ x \in c \wedge (\exists j \in J\colon\ x \in j)\}$.

Lemma 4. *For any interval c and any natural number k, $\mu(\Delta \cap c) \leqslant \mu(\Delta_k \cap c) + r_k$.*

Proof. Since $\Delta_k \subseteq \Delta$, for any natural k we have that $\Delta \cap c \subseteq (\Delta_k \cap c) \cup (\Delta - \Delta_k)$. Taking measure we obtain $\mu(\Delta \cap c) \leqslant \mu(\Delta_k \cap c) + \mu(\Delta - \Delta_k)$. By Proposition 2 we have $\mu(\Delta \cap c) \leqslant \mu(\Delta_k \cap c) + r_k$. □

Lemma 5. *For any interval c any natural numbers k and l such that $k \leqslant l$, $\mu(\Delta_l \cap c) \leqslant \mu(\Delta_k \cap c) + r_k - r_l$.*

Proof. Obvious, from Proposition 3 and $\Delta_l \cap c \subseteq (\Delta_k \cap c) \cup (\Delta_l - \Delta_k)$. □

Lemma 6. *$\mu(\Delta_k)$ is computable for any k, and $\mu(\Delta_k \cap c)$ is computable for any k and for any interval $c = (a, b)$ where a and b are rationals.*

Proof. Δ_k is a finite set of known intervals with rationals end points. An algorithm for the measure of Δ_k and for the measure of $\Delta_k \cap c$ can be easily given. □

3.1. Determination of the first digit

We will compute b_1. We divide the interval $[0, 1]$ in two halves $c_0^1 = [0, \frac{1}{2}]$ and $c_1^1 = [\frac{1}{2}, 1]$. We will have to determine a suitable index p_1. It is clear that

$$\mu(\Delta_{p_1} \cap c_0^1) + \mu(\Delta_{p_1} \cap c_1^1) = \mu(\Delta_{p_1}) \leqslant s_{p_1}.$$

Adding $r_{p_1} + r_{p_1}$ on each side of this inequality and using the definition of r_{p_1} gives

$$(\mu(\Delta_{p_1} \cap c_0^1) + r_{p_1}) + (\mu(\Delta_{p_1} \cap c_1^1) + r_{p_1}) \leqslant s_{p_1} + r_{p_1} + r_{p_1} = s + r_{p_1} < \varepsilon + r_{p_1}.$$

It is impossible that both terms $\mu(\Delta_{p_1} \cap c_0^1) + r_{p_1}$ and $\mu(\Delta_{p_1} \cap c_1^1) + r_{p_1}$ be greater or equal to $(\varepsilon + r_{p_1})/2$. If they were, we would have

$$\varepsilon + r_{p_1} = \frac{\varepsilon + r_{p_1}}{2} + \frac{\varepsilon + r_{p_1}}{2} \leqslant (\mu(\Delta_{p_1} \cap c_0^1) + r_{p_1}) + (\mu(\Delta_{p_1} \cap c_1^1) + r_{p_1})$$
$$< \varepsilon + r_{p_1}$$

a contradiction. Thus, the following proposition is true

$$\left(\mu(\Delta_{p_1} \cap c_0^1) < \frac{\varepsilon + r_{p_1}}{2} - r_{p_1}\right) \vee \left(\mu(\Delta_{p_1} \cap c_1^1) < \frac{\varepsilon + r_{p_1}}{2} - r_{p_1}\right).$$

Now we determine the value of p_1. It has to be large enough so that the error r_{p_1} is sufficiently small to guarantee that even if all the remaining intervals that have not yet been enumerated in step p_1 fall in $c_{b_1}^1$, the whole $c_{b_1}^1$ will not be completely covered by Δ. We need $\mu(\Delta_{p_1} \cap c_{b_1}^1) + r_{p_1} < \frac{1}{2}$. We know that the measure $\mu(\Delta_{p_1} \cap c_{b_1}^1) + r_{p_1} < (\varepsilon + r_{p_1})/2$. Theorem 1 states that $r_{p_1} < 5\varepsilon/2p_1$. So, for $p_1 = 5$ we obtain $r_{p_1} < \varepsilon/2$. Then, $\mu(\Delta_{p_1} \cap c_{b_1}^1) + r_{p_1} < (\varepsilon + r_{p_1})/2 = \varepsilon/2 + \varepsilon/4 < \varepsilon \leqslant 1/2$. Using Lemma 4 we obtain $\mu(\Delta \cap c_{b_1}^1) < \frac{1}{2} = \mu(c_{b_1}^1)$. This means that the union of all the intervals belonging to Δ will never cover the whole interval $c_{b_1}^1$, whose measure is $\frac{1}{2}$. Thus, there exist real numbers belonging to no interval of Δ that fall in the interval $c_{b_1}^1$. These have their first digit equal to b_1. We define

$$b_1 = \begin{cases} 0 & \text{if } \mu(\Delta_{p_1} \cap c_0^1) \leqslant \mu(\Delta_{p_1} \cap c_1^1), \\ 1 & \text{otherwise.} \end{cases}$$

3.2. *Determination of the nth digit, for $n > 1$*

Let us assume that we have already computed $b_1, b_2, \ldots, b_{n-1}$ and that at each step m, $1 \leqslant m < n$,

$$\mu(\Delta_{p_m} \cap c_{b_m}^m) + r_{p_m} < \frac{1}{2^m}\left(\varepsilon + \sum_{j=1}^{m} 2^{j-1} \cdot r_{p_j}\right)$$

and we have chosen $p_m = 5 \times 2^{2m-2}$. We have to prove that this condition holds for $m = n$ and $p_n = 5 \times 2^{2n-2}$, and that we can computably determine b_n. We split the interval $c_{b_{n-1}}^{n-1}$ in two halves of measure $1/2^n$, $c_0^n = [0.b_1 b_2 \ldots b_{n-1},\ 0.b_1 b_2 \ldots b_{n-1} 1]$ and $c_1^n = [0.b_1 b_2 \ldots b_{n-1} 1, 0.b_1 b_2 \ldots b_{n-1} 111111\ldots]$. As they cover the interval $c_{b_{n-1}}^{n-1}$, we have

$$\mu(\Delta_{p_n} \cap c_0^n) + \mu(\Delta_{p_n} \cap c_1^n) = \mu(\Delta_{p_n} \cap c_{b_{n-1}}^{n-1}).$$

Since $p_n \geqslant p_{n-1}$ and using Lemma 5 we obtain

$$\mu(\Delta_{p_n} \cap c_0^n) + \mu(\Delta_{p_n} \cap c_1^n) \leqslant \mu(\Delta_{p_{n-1}} \cap c_{b_{n-1}}^{n-1}) + r_{p_{n-1}} - r_{p_n}.$$

Adding $r_{p_n} + r_{p_n}$ to both sides of this inequality we obtain

$$(\mu(\Delta_{p_n} \cap c_0^n) + r_{p_n}) + (\mu(\Delta_{p_n} \cap c_1^n) + r_{p_n}) \leqslant \mu(\Delta_{p_{n-1}} \cap c_{p_{n-1}}^{n-1}) + r_{p_{n-1}} + r_{p_n}$$

and by the previous equation

$$(\mu(\Delta_{p_n} \cap c_0^n) + r_{p_n}) + (\mu(\Delta_{p_n} \cap c_1^n) + r_{p_n}) < \frac{1}{2^{n-1}} \left(\varepsilon + \sum_{j=1}^{n} 2^{j-1} \cdot r_{p_j} \right).$$

Hence, one of the two terms, $\mu(\Delta_{p_n} \cap c_0^n) + r_{p_n}$ or $\mu(\Delta_{p_n} \cap c_1^n) + r_{p_n}$, must be less than

$$\tfrac{1}{2^n}(\varepsilon + \sum_{j=1}^{n} 2^{j-1} \cdot r_{p_j}).$$

This means that the following proposition is true:

$$\left(\mu(\Delta_{p_n} \cap c_0^n) < \frac{\varepsilon + \sum_{j=1}^{n} 2^{j-1} \cdot r_{p_j}}{2^n} - r_{p_n} \right)$$
$$\vee \left(\mu(\Delta_{p_n} \cap c_1^n) < \frac{\varepsilon + \sum_{j=1}^{n} 2^{j-1} \cdot r_{p_j}}{2^n} - r_{p_n} \right).$$

We define b_n as the first index i such that the interval c_i^n is less covered by Δ_{p_n},

$$b_n = \begin{cases} 0 \text{ if } \mu(\Delta_{p_n} \cap c_0^n) \leqslant \mu(\Delta_{p_n} \cap c_1^n), \\ 1 \text{ otherwise.} \end{cases}$$

By Theorem 1 we have

$$\sum_{j=1}^{n} 2^{j-1} \cdot r_{p_j} < \varepsilon \cdot \sum_{j=1}^{n} \frac{2^{j-1}}{2^{2j-1}} = \varepsilon \cdot \sum_{j=1}^{n} 2^{-j} < \varepsilon.$$

From the last inequality and from the definition of b_n we obtain

$$\mu(\Delta_{p_n} \cap c_{b_n}^n) + r_{p_n} < \frac{1}{2^n} \left(\varepsilon + \sum_{j=1}^{n} 2^{j-1} \cdot r_{p_j} \right) < \frac{2\varepsilon}{2^n} \leqslant \frac{1}{2^n}$$

and using Lemma 4 we deduce $\mu(\Delta \cap c_{b_n}^n) < 1/2^n = \mu(c_{b_n}^n)$. Hence, the set Δ does not cover the interval $c_{b_n}^n$. There must be real numbers in the interval $c_{b_n}^n$ that belong to no interval of Δ.

Theorem 7. *The number v is computable and absolutely normal.*

Proof. In our construction we need only to compute the measure of the sets $(\Delta_{p_n} \cap c_{b_n}^n)$. Then, by Lemma 6, v is computable.

Let us prove that v is external to every interval of Δ. Suppose not. Then, there must be an open interval $I \in \Delta$ such that $v \in I$. Consider the intervals $c_{b_1}^1, c_{b_2}^2, c_{b_3}^2, \ldots$

By our construction, v belongs to every $c^n_{b_n}$. Let us call c the first interval $c^n_{b_n}$ of the sequence such that $c^n_{b_n} \subset I$. Such an interval exists because the measure of $c^n_{b_n}$ goes to 0 as n increases. But then the interval c is covered by Δ. This contradicts that in our construction at each step n we choose an interval $c^n_{b_n}$ not fully covered by Δ. Thus, v belongs to no interval of Δ, so by Sierpinski's result, v is absolutely normal. □

Finally, it follows from Theorem 1 that the bound s on $\mu(\Delta)$ is also computable.

Corollary 8. *The real number s is computable.*

Proof. Let us define the sequence of rationals in base 2, $a_n = s_{5\varepsilon \cdot 2^{n-1}}$. By Theorem 1 we have $|s - a_n| = s - s_{5\varepsilon \cdot 2^{n-1}} = r_{5\varepsilon \cdot 2^{n-1}} < 2^{-n}$. Then, it is possible to approximate s by a computable sequence of rationals a_n such that the first n digits of a_n coincide with the first n digits of s. Therefore, s is computable. □

3.3. About Sierpinski's ξ number

We finish this section with some observations about the absolutely normal number defined by Sierpinski, ξ the first real number external to $\Delta(1)$, that is, $\xi = \min(E(1))$. As we see, Sierpinski defines ξ fixing $\varepsilon = 1$. Since our construction requires $\varepsilon \in (0, \frac{1}{2}]$, we do not obtain the number ξ. Under a slight modification of our construction we can allow ε to be any computable number in the interval $(0, 1)$, by defining $p_n = \lfloor 5 \cdot 2^{2n-2}\varepsilon/1 - \varepsilon\rfloor + 1$.

Now we can speak of the family of numbers that are definable using Sierpinski's notion for different values of ε. Fix ε to be any computable real in $(0, 1)$ and let $\xi = \min(E(\varepsilon))$. Then ξ is definable in our construction, in binary notation, in the following way:

$$b_n = \begin{cases} 0 & \text{if } \mu(\Delta_{p_n} \cap c^n_0) < \frac{1}{2^n}(\varepsilon + \sum_{j=1}^{n} 2^{j-1} \cdot r_{p_j}) - r_{p_n}, \\ 1 & \text{otherwise.} \end{cases}$$

For each n, ξ either falls in c^n_0 or in c^n_1; therefore, we get a determination of each digit of its binary expansion. If we could prove that

$$\frac{1}{2^n}(\varepsilon + \sum_{j=1}^{n} 2^{j-1} \cdot r_{p_j}) - r_{p_n}$$

is irrational and computable, then we could assert that ξ is computable. Without this assumption we can assert a weaker property: ξ is *computably enumerable*.

A real number is computably enumerable if there is a computable non decreasing sequence of rationals which converges to that number. Every computable number is computably enumerable, but the converse is not true. It is possible that a real number r be approximated from below by a computable non decreasing sequence of rationals but that there is no function which effectively gives each of the fractional digits of r.

This happens when it is not possible to bound the error of approximating the number by any computable sequence of rationals.

Let us sketch the proof that ξ is computably enumerable, for any computable real $\varepsilon \in (0,1]$. Since $\Delta(\varepsilon)$ is a recursively enumerable set of intervals, we can scan them one by one. At each step we single out the first rational in $[0,1]$ which is external to all the intervals scanned up to the moment. This procedure is computable and determines a non decreasing sequence of rationals which converges to ξ. This proves that ξ is computably enumerable. However, because of the shape of the intervals in $\Delta(\varepsilon)$, it does not seem easy to bound the error of this method of approximating ξ.

4. Other computable absolutely normal numbers

The construction we gave defines ν, a computable absolutely normal number in base 2. We can adapt the construction to define numbers in any other bases: To compute a number in a base $q \geqslant 2$, at each step we should divide the interval selected in the previous step in q parts. In the nth step we determine the nth digit defining the intervals $c_0^n, c_1^n, \ldots, c_{q-1}^n$ (of measure $1/q^n$), where

$$c_i^n = [i/q^n \sum_{j=1}^{n-1} b_j/q^j,\ i + 1/q^n + \sum_{j=1}^{n-1} b_j/q^j] \text{ for } 0 \leqslant i \leqslant q-1.$$

We will choose $p_n = 5 \cdot (q-1) \cdot 2^{2n-2}$ and following the same steps as in the construction of a number in base 2, there must be an index i such that

$$\mu(\Delta_{p_n} \cap c_i^n) < \frac{1}{q^n} \left(\varepsilon + (q-1) \sum_{j=1}^{n} 2^{j-1} \cdot r_{p_j} \right) - r_{p_n}.$$

As before, we define b_n as the first index corresponding to the interval least covered by Δ_{p_n},

$$b_n = \min_{0 \leqslant i \leqslant q-1} \{i\colon\ (\forall j\colon\ 0 \leqslant j \leqslant q-1\colon\ \mu(\Delta_{p_n} \cap c_i^n) \leqslant \mu(\Delta_{p_n} \cap c_j^n))\}.$$

In principle, for different bases the numbers will be distinct (they will not be ν expressed in different bases), while they will all be examples of computable absolutely normal numbers.

The definition of absolute normality is asymptotic, that is, it states a property that has to be true in the limit. Thus, given an absolutely normal number, we can alter it by adding or removing a finite number of digits of its fractional expansion to obtain an absolutely normal number. For example, we could fix an arbitrary number of digits of the fractional expansion and complete the rest with the digits of ν. However, we wonder whether it is possible to obtain an absolutely normal number by fixing a priori infinitely many digits and filling in the free slots. Obviously with only finitely many free slots we may not obtain an absolutely normal number (for example, fix all the digits to be 0 except finitely many). Likewise we may not obtain an absolutely normal number by fixing infinitely many digits and leaving free also infinitely many slots: if we fix 0 in the even positions we will never obtain an absolutely normal number because we will never find the string “11” in the fractional expansion. The same thing happens if we fix the digits in the positions which are multiples of 3, 4, etc. It is clear that the possibility to obtain an absolutely normal number depends on the values we assign to the fixed positions (if we set the even positions with the digits of the even positions of v it would be trivial to complete the rest to obtain an absolutely normal number). Finally, we wonder what happens if the digits we fix are progressively far apart, for example in the positions which are powers of 2 or in the positions which are Fibonacci numbers.

Let us also note that absolute normality is invariant under permutation of digits.

Acknowledgements

The present work was originated in a discussion with Guillermo Martinez, who pointed out that Sierpinski gave the first example of an absolutely normal number in 1916. The authors are also indebted to Cristian Calude and Max Dickmann for their valuable comments.

References

[1] D.H. Bailey, R.E. Crandall, On the random character of fundamental constant expansions, http://www.perfsci.com/free/techpapers/freepapers.html, 2000.
[2] E. Borel, Les probabilités dénombrables et leurs applications arithmétiques, Rend. Circ. Mat. Palermo 27 (1909) 247–271.
[3] G.J. Chaitin, A theory of program size formally identical to information theory, ACM 22 (1975) 329–340.
[4] G. Martin. Absolutely abnormal numbers, Amer. Math. Monthly, to appear.
[5] M.W. Sierpinski, Démonstration élémentaire du théorème de M. Borel sur les nombres absolument normaux et détermination effective d’un tel nombre, Bull. Soc. Math. France 45 (1917) 127–132.
[6] M.W. Sierpinski, Elementary Theory of Numbers, Polska Akademia Nauk Monografie Mathematyczne, 44, Warszawa, 1964.
[7] A. Turing, A note on normal numbers, in: J.L. Britton (Ed.), Collected Works of A. M. Turing, Pure Mathematics, North-Holland, Amsterdam, 1992, pp. 117–119.